科学数据开放共享障碍的实证研究*

■ 盛小平[1] 吴红[2] 胡冰洁[2]
[1] 上海大学图书情报与档案系 上海 200444 [2] 华南师范大学经济与管理学院 广州 510006

摘要：［目的/意义］归纳与验证科学数据开放共享障碍的观点，以促进科学数据开放共享的理论研究。［方法/过程］基于科学数据开放共享障碍的理论分析，提炼出64个研究假设，并以此作为调查问卷的问项指标，运用问卷调查和数据处理与分析，对相关研究假设进行验证。［结果/结论］科学数据开放共享的障碍主要包括26种管理与实施障碍、13种法律与资金障碍、9种技术障碍、8种认知障碍、4种成本障碍、4种用户障碍。我们应该针对上述障碍制定相关解决方案或对策，以着实提升我国科学数据开放共享水平。

关键词：科学数据开放 共享障碍 实证分析
分类号：G203
DOI:10.13266/j.issn.0252-3116.2019.17.003

1 引言

随着开放获取(Open Access)、开放科学(Open Science)运动的快速发展，科学数据开放共享成为开放领域的一个新议题。尽管科学数据开放共享实践已经取得了一定的进展，比如，许多组织(如经济合作与发展组织、欧洲研究理事会、英国研究理事会、惠康基金会、美国国家科学基金会、美国科学与技术理事会、加拿大社会科学与人文研究理事会、爱尔兰科学基金会等)制定了科学数据开放共享政策[1]，建立了一些包括联合国开放科学数据库、经济合作与发展组织图书馆开放科学数据、国际货币基金组织数据库、非洲开放数据在内的大型开放科学数据资源库[2]和一些包括如开放科学中心、生物医学中心、公共医学中心、中国地球系统科学数据共享平台、中国人口与健康科学数据共享平台等在内的科学数据开放共享平台。但是，科学数据开放共享目前仍面临许多问题和障碍，如法律缺失、侵犯隐私、信息技术基础设施不完善、学术报酬机制不健全[3-4]，以及认知障碍、治理障碍、成本障碍、技能障碍、伦理障碍、社会-技术障碍等[5-8]。这里选取"科学数据"(Scientific Data)或"研究数据"(Research Data)、"开放数据"(Open Data)作为关键词，并把"开放获取"(Open Access)或"共享"(Sharing)、"障碍"(Barriers或Impediments)作为并列关键词，检索中国知网和Elsevier、Springer、EBSCO、Emerald、Taylor & Francis、SAGE Premier、Wiley、Proquest等全文数据库以及网络文献，结果发现尽管已有文献对科学数据开放共享障碍进行了理论[9-11]与调查分析[12]，但鲜见实证研究成果。于是，本文拟在对科学数据开放共享障碍相关观点进行理论归纳的基础上，提出相关研究假设，然后利用调查问卷进一步验证相关假设，以便证实存在哪些科学数据开放共享障碍，为今后人们克服相关障碍并促进科学数据开放共享实践提供参考。

2 科学数据开放共享障碍研究假设的提出

从现有研究成果来看，人们对科学数据开放共享障碍类型有如下多种不同观点：①3类障碍：技术障碍、行为和组织机构障碍、法律障碍[9]。②6类障碍：机构障碍、任务复杂性障碍、使用与参与障碍、法律障碍、信息质量障碍、技术障碍[7]；或者是技术障碍、动机障碍、经济障碍、政治障碍、法律障碍、伦理障碍[11]；或者是经济障碍、技术障碍、法律障碍、文化障碍、管理障碍、风险相关障碍[8]。③7类障碍：治理障碍、经济问题、许可和法律框架障碍、数据特性、元数据、访问障碍、技能障碍[6]。④10类障碍：可用性与访问障碍、发现能

* 本文系国家社会科学基金项目"开放科学环境下的科学数据开放共享机制与对策研究"(项目编号：18ATQ007)研究成果之一。
作者简介：盛小平(ORCID:0000-0002-6341-6973)，教授，博士，博士生导师，E-mail:shengxp68@126.com；吴红(ORCID:0000-0003-2239-9561)，硕士研究生；胡冰洁(ORCID:0000-0002-7202-5341)，硕士研究生。
收稿日期：2018-10-12 修回日期：2019-04-02 本文起止页码：23-30 本文责任编辑：易飞

力障碍、适用性障碍、理解能力障碍、质量障碍、关联和组合数据障碍、可比性和兼容性、元数据、与数据提供者的交互障碍、开放和上传障碍[13]。综合现有观点，本文将从认知、经济、法律、技术、管理、实施 6 个方面来解析科学数据开放共享障碍。

2.1 认知障碍及其研究假设

科学数据开放共享的认知障碍主要是指人们对科学数据开放共享存在某些错误认识或忧虑而不愿参与实践的各种现象与原因。这主要包括：①各种忧虑，如担心研究观点受到挑战或反驳，担心破坏原始研究计划，担心受到援助请求的困扰，担心无法保证数据的质量和数量，担心降低自身的竞争优势[4]；担心受到批评，担心失去对数据的控制，担心研究数据被他人抢先发布，担心数据误用[11,14-15]；担心泄露数据中的隐私信息，害怕失去出版机会[16]。②认知缺失，如缺乏获取他人研究数据的意识[17]；缺乏对数据潜在用途的认识[9]；不了解调查数据的来源[18]；缺乏理解和合理利用科学数据的知识，缺乏挖掘数据价值的知识，缺乏统计知识[7]；缺乏创建和保存元数据的知识[19]。基于这些观点，可提出以下 12 个科学数据开放共享认知障碍的研究假设：A1：人们缺少共享科学数据的意识；A2：人们担心降低自身的竞争优势；A3：研究者担心科学数据被别人误解；A4：研究者担心科学数据被别人误用；A5：研究者担心存在研究错误而受到批评；A6：研究者担心打乱研究计划；A7：研究者担心受到别人干扰；A8：研究者担心失去对数据的控制；A9：研究者担心泄露数据中的隐私信息；A10：研究者害怕失去出版机会；A11：数据用户缺少合理利用科学数据的知识；A12：数据用户缺乏创建和保存元数据的知识。

2.2 经济障碍及其研究假设

科学数据开放共享的经济障碍是指科学数据开放共享在成本问题（包括时间、人力和资金成本）或收益损失方面可能遇到的问题或障碍。这主要包括：①费用与成本，如搜集、提供、存储和传递数据的时间成本高，提供后续研究需要的时间和人力资本高，使用共享数据的时间和人力成本高[4,7,15,20]。②将数据收入作为财政收入的一部分而不愿共享数据[7]。③存在可能的或潜在的经济损失[11]。④缺乏处理和出版数据的资金[9]。基于上述认识，可提出以下 8 个科学数据开放共享经济障碍的研究假设：B1：搜集、提交和存储科学数据的时间成本高；B2：搜集、提交和存储科学数据的人力成本高；B3：提供、传递、维护和管理科学数据的时间成本高；B4：提供、传递、维护和管理科学数据的人力成本高；B5：缺乏共享科学数据的报酬或奖励机制；B6：缺乏足够的科学数据开放共享资金投入；B7：需要增加预算外的研究成本；B8：因科学数据开放共享而存在的潜在经济损失。

2.3 法律障碍及其研究假设

科学数据开放共享的法律障碍是指科学数据开放共享在知识产权、个人隐私、数据安全与保护等方面可能遇到的问题或障碍。这主要包括：①知识产权问题，如存在知识产权的控制[15]，难以获取科学数据开放共享的许可[9]，知识产权保护机制不完善[21]，未解决多作者数据共有权问题[22]；面临诉讼的风险[9]。②隐私保护问题[11]，如数据共享易侵犯受访者隐私[4]；数据提供者要求使用者去除隐私数据[15]；参与者隐私信息保护问题未解决[20]；法律保护的数据隐私范围较大[9]。③数据安全问题，如因考虑国家安全而不能公开数据[14]，或未解决数据安全问题[7]。④法律不健全问题，如缺乏科学数据开放共享的法律框架，数据共享和数据使用权的法律条款相互矛盾[9]；现有的知识产权和商业秘密法律法规不利于鼓励开放科学实践[10]；研究者无权利用开放数据[19]。基于上述认识，可提出以下 7 个科学数据开放共享法律障碍的研究假设：C1：缺乏完善的科学数据开放共享法律体系；C2：存在科学数据开放共享安全问题；C3：存在科学数据开放共享的隐私问题；C4：容易失去对科学数据所有权和版权的控制；C5：存在科学数据的知识产权问题；C6：存在科学数据的许可问题；C7：可能面临科学数据开放共享的争议和诉讼风险。

2.4 技术障碍及其研究假设

科学数据开放共享的技术障碍是指科学数据开放共享过程中所遇到的与数据提交、存储、分享、使用和维护相关的技术与设施问题或障碍。这主要包括：①数据质量与软件问题，如数据准确性与及时性差，没有关于数据质量的描述信息，缺少元数据[9]或没有共同商定的元数据[13]；数据不匹配[23]；数据软件包不兼容，提供的数据格式不标准，新旧数据格式不兼容，数据不完整或过于复杂[11,14]；缺少处理开放数据的标准软件[7]。②缺少标准，如数据集的保存和维护缺乏标准化[4]；缺乏元数据标准[7]。③数据访问障碍，如访问数据很困难[6]；某些平台需要用户注册和登录才能访问数据[6]；缺少共享数据的公共数据库[22]；存在语言障碍[9]。④数据处理障碍，如缺乏处理大规模数据集的技术[22]；缺乏数据保护技术，缺乏保证数据质量、数据标准化的技术[18]；缺乏有效的的技术解决方案[11]。

⑤缺少数据基础设施或辅助工具,如缺乏用来共享数据的基础设施[9];缺少数据共享的辅助工具和系统帮助,缺少能准确找到所需科学数据的索引或其他搜索方式,缺乏科学数据开放共享平台[7]。基于上述认识,可提出以下 11 个科学数据开放共享技术障碍的研究假设:D1:存在科学数据完整性较差的问题;D2:存在科学数据准确性较差的问题;D3:存在科学数据及时性较差的问题;D4:存在科学数据格式不标准问题;D5:存在数据软件不兼容的问题;D6:缺乏科学数据开放共享平台或系统;D7:缺少科学数据共享的辅助工具(如索引、搜索引擎);D8:缺乏元数据;D9:存在国家间共享科学数据的语言障碍;D10:缺乏科学数据保护技术;D11:缺乏科学数据开放共享标准。

2.5 管理障碍及其研究假设

科学数据开放共享的管理障碍是指个人或组织在科学数据开放共享活动中遇到的各种管理问题或障碍。这主要包括:①领导与政策缺失,如缺少科学数据开放共享的领导[8];缺乏数据共享管理和协调机制[21];缺乏有效统一的科学数据开放共享政策[24]。②程序障碍,如数据管理程序繁琐[4],数据管理措施混乱[7]。③数据发布障碍,如缺乏统一的数据公布政策,缺乏数据公布的途径[7]。④数据输入障碍,如缺少对用户输入过程的数据管理,无法保证用户输入数据的质量[7];缺乏保障元数据质量的机制[6]。⑤缺少激励,如缺乏有效的数据开放激励措施[7],数据生产者和管理者之间缺乏互惠机制[11]。⑥缺少信任[11],如机构存在信息封锁和信息壁垒现象[25];缺少开放共享的文化[7]。基于上述分析,可提出以下 14 个科学数据开放共享管理障碍的研究假设:E1:缺乏有效的科学数据开放共享政策;E2:缺乏科学数据开放共享协调机制;E3:缺乏科学数据开放共享管理机制;E4:存在与科学数据开放共享相冲突的管理制度;E5:存在限制科学数据开放共享的商业合同;E6:缺少有效的科学数据开放共享计划;E7:缺少科学数据开放共享途径;E8:缺少与数据用户交互的经验;E9:缺少处理用户反馈信息的环节;E10:缺少共享数据的激励机制;E11:缺少开放共享的组织文化;E12:科学数据管理措施混乱;E13:科学数据管理程序复杂;E14:数据生产者和管理者之间缺乏互惠机制。

2.6 实施障碍及其研究假设

科学数据开放共享的实施障碍是指科学数据开放共享活动实施过程中所遇到的各种操作性问题或障碍。这主要包括:①数据获取障碍,如数据提供者限制数据的获取[15];缺乏理解数据的能力[9];缺乏判断科学数据价值的能力[26];缺乏识别科学数据开放共享的能力[5];缺少使用数据的能力[13]。②数据关联障碍,如连接数据困难,数据不能与其他数据连接起来[13]。③存在数据结果冲突问题,如同一数据的研究结果可能相互矛盾,存储在不同系统中的相似数据产生不同的结果[7]。④缺乏科学数据开放共享指南[11];上传数据困难[13]。⑤缺少合作,如各类机构之间缺乏科学数据开放共享合作,不同机构间存在知识共享的困难[9]。⑥缺少数据资源开放共享领域的专业人才[24]。基于上述分析,可提出以下 12 个科学数据开放共享实施障碍的研究假设:F1:数据提供者与数据用户之间缺乏信任;F2:缺乏科学数据开放共享指南;F3:数据提供者限制了数据的使用;F4:机构内部或机构之间缺乏科学数据开放共享的各种合作;F5:科学数据市场调节失灵;F6:科学数据开放共享各国发展不平衡;F7:开放共享科学数据在现有科学数据中所占比例太小;F8:科学数据被垄断;F9:缺乏数据资源开放共享领域的专业人才;F10:数据用户缺乏判断或理解科学数据价值的能力;F11:数据用户缺乏数据获取能力;F12:数据用户缺乏数据处理能力。

3 问卷调查

3.1 先期调查

在 2017 年 12 月 8－9 日,笔者利用在湖南某所高校进行学术交流的机会,面向该校部分老师和中级以上或取得硕士学位以上的图书馆员,发放了 50 份题为“科学数据开放共享利益、障碍与现状的问卷调查”,并采访了 6 位专家。由于先期调查涉及科学数据开放共享的利益、障碍与现状三个主题,调查项目总数较多,但每个主题调查问题相对不全,因此,先期调查得到的数据不能很好地支撑研究。于是,在汲取专家建议基础上,笔者于 2018 年 4 月修订了原调查问卷,分别设计了“科学数据开放共享障碍的调查问卷”和“科学数据开放共享利益的调查问卷”。

3.2 问卷修订与调查

以上述 64 个研究假设(A1,A2,A3……F12)重新设计了“科学数据开放共享障碍的调查问卷”,采用五分制等级评估标准——“非常同意”“同意”“不确定”“不同意”“非常不同意”,分别对应五分制李克特量表中的 5、4、3、2、1。在 2018 年 5 月 23 日至 6 月 30 日,向广州大学城内 6 所高校的师生发放纸质调查问卷 200 份,实际回收 132 份;通过问卷星邀请国内其他地区高

校教师和研究生(均不限制专业)参与本次问卷调查,回收 128 份答卷,由此共得到 260 份调查问卷,去除无效答卷后得到有效问卷 252 份。其中,去除无效答卷的标准是:①问卷填答个别题目缺漏,有 1 题或 1 题以上未作答的情况;②出现明显错填、乱填的情况。

调查由现场调查和网络调查两部分组成。在现场纸质调查中,调查者先通过当面咨询并了解受访对象身份后,邀请其接受问卷调查;在网络调查中,被调查者是笔者熟知的教师或研究生以及这些教师或研究生了解的其他人。由于这些调查对象经常使用科学数据,进行科学研究,其答卷有一定的代表性。

4 实证分析

下面运用 SPSS19.0 统计分析软件,从调查问卷的效度分析、信度分析、因子分析等方面来对科学数据开放共享障碍进行实证分析。

4.1 问卷的效度与信度分析

利用所有题项(即 A1、A2、A3、……F12)与其总分的相关性分析来评判调查问卷的效度。运行SPSS19.0后,结果显示所有题项与其总分的相关系数在 0.176 ~0.787 之间,且显著性(双侧)都为 0.000,在 0.01 水平(双侧)上显著相关,由此说明调查问卷具有较高效度,适合于作实证分析。同时,运用克隆巴赫 α(Cronbach' sAlpha)系数来检验调查问卷的信度。运行SPSS19.0后,结果显示 64 个题项的克隆巴赫 α 值为 0.981,远高于基准值 0.800,说明整份调查问卷满足信度检测要求。

4.2 第一次与第二次因子分析和因子信度分析

从表 1 所示的 KMO 和 Bartlett 的检验来看,KMO 的值为 0.954,显著性概率 P 值为 0.000 <0.05,即说明 A1、A2…F12 等 64 个变量间具有较高的共同因子,适合做因子分析。

表 1 KMO 和 Bartlett 的检验

取样足够的 Kaiser-Meyer-Olkin 度量		0.954
Bartlett 的球形度检验	近似卡方	13538.024
	自由度	2016
显著性	0.000	

运用主成分分析法,并萃取特征值大于 1 的共同因子,运行 SPSS19.0 软件后,获得 10 个共同因子,但是由于这些因子的题项较为杂乱,个别题项的因子负载量小于 0.400,因此,该方法不太适宜做因子分析。于是,改用设定因子数量的方法来提取因子。本文将萃取的因子个数设定为 6 个,并采用主成分分析提取方法,再次运行 SPSS19.0 软件,得知 6 个因子可以解释的总方差为 62.518%(见表 2),满足大于 50.0% 的要求,且在 6 个因子的旋转成分矩阵中,每个因子包含的负荷量均大于 0.400,说明各题项与所在因子的相关性满足要求。第一次 6 个因子信度分析(见表 3)显示,每个因子的克隆巴赫 α 系数都在 0.797 以上,由此说明 6 个因子的各个变量具有较高的可信度。不过,因子 2 中的题项 A2 对应的"题项已删除的克隆巴赫 α 系数"为 0.948 >0.947,不具备可靠性,宜删除。

表 2 解释的总方差

成分	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的%	累计%	合计	方差的%	累计%	合计	方差的%	累计%
1	29.398	45.935	45.935	29.398	45.935	45.935	12.313	19.240	19.240
2	2.943	4.598	50.533	2.943	4.598	50.533	8.990	14.047	33.287
3	2.774	4.334	54.866	2.774	4.334	54.866	6.389	9.983	43.270
4	1.998	3.122	57.988	1.998	3.122	57.988	5.198	8.121	51.392
5	1.485	2.320	60.308	1.485	2.320	60.308	4.192	6.550	57.942
6	1.415	2.210	62.518	1.415	2.210	62.518	2.929	4.577	62.518
7	1.261	1.970	64.488						
…	…	…	…						
64	0.058	0.090	100.000						

表 3 第一次因子信度分析

信度指标 \ 因子	因子 1	因子 2	因子 3	因子 4	因子 5	因子 6
克隆巴赫 α 系数	0.968	0.947	0.921	0.876	0.914	0.797
项目个数	26	13	9	8	4	4

在删除了题项 A2 后,运用同样的因子分析方法对余下的 63 个题项进行第二轮因子分析和因子信度分析。运行 SPSS 19.0 后,得知 KMO 的值为 0.954,显著性概率 P 值为 0.000 <0.05,满足因子分析要求。新的

盛小平，吴红，胡冰洁.科学数据开放共享障碍的实证研究[J].图书情报工作,2019,63(17):23－30.

6个因子解释总方差是62.825%，但是，因子2的题项A1的"题项已删除的克隆巴赫α系数"为0.949，大于该因子2的克隆巴赫α系数0.948，因此，题项A1也不具备可靠性，宜删除。

4.3 第三次因子分析和因子信度分析

在删除A1后，对余下的62个题项进行第三次因子分析和因子信度分析。运行SPSS 19.0后，得知KMO的值为0.955，显著性概率P值为0.000＜0.05，满足因子分析要求。然后，采用主成份分析提取方法和具有Kaiser标准化的正交旋转法，并将萃取的因子个数设定为6个，运行SPSS 19.0后发现，旋转在10次迭代后收敛，且新的6个因子的解释总方差是63.165%，满足大于50%的要求。而从6个因子旋转成分矩阵（见表4）来看，6个因子的负荷量均大于0.400，各因子包含的题项数量均大于3。其中，因子1由D10、D11、E1-E14、F1-F10题项组成，共同反映出科学数据开放共享面临的管理与实施问题，该因子可命名为"管理与实施障碍"；因子2由B5-B8、C1-C7题项组成，既反映出科学数据开放共享面临的经济损失问题，也反映了科学数据开放共享面临的法律问题，该因子可命名为"法律与资金障碍"；因子3由D1-D9题项组成，共同反映出科学数据开放共享面临的技术问题，该因子可命名为"技术障碍"；因子4由A3-A10题项组成，共同反映出科学数据开放共享面临的认知缺陷，该因子可命名为"认知障碍"；因子5由B1-B4题项组成，共同反映出科学数据开放共享面临的成本问题，该因子可命名为"成本障碍"；因子6由A11、A12、F11、F12题项组成，共同反映出用户在科学数据开放共享方面面临的问题，该因子可命名为"用户障碍"。

第三次6个因子信度分析（见表5）显示，每个因子克隆巴赫α系数均在0.797以上，由此说明6个因子的各个变量具有较高可信度。而且，每个因子中各题项对应的"校正题项总计相关性"都在0.400以上，各题项对应的"题项已删除的克隆巴赫α系数"都没有超过其克隆巴赫α系数，由此说明6个因子信度分析满足要求，也就说明，原64个研究假设除A1和A2两个题项不成立外，其余62个研究假设都得到验证，即命题是成立的。

5 结论与建议

上述研究结果验证了科学数据开放共享存在管理与实施障碍、法律与资金障碍、技术障碍、认知障碍、成本障碍、用户障碍6类主要障碍（见表6）。

表4 旋转成分矩阵（第三次因子分析）

题项	因子成分					
	管理与实施障碍	法律与经济损失障碍	技术障碍	认知障碍	成本障碍	用户障碍
E12	0.707	0.091	0.266	0.238	0.108	0.186
E6	0.685	0.284	0.196	0.186	0.171	－0.063
F4	0.653	0.215	0.038	0.148	0.235	0.203
E7	0.646	0.266	0.243	0.144	0.083	－0.013
E14	0.645	0.301	0.141	0.161	0.027	0.159
E11	0.641	0.129	0.218	0.208	0.182	0.138
E13	0.638	0.171	0.235	0.121	0.136	0.278
F1	0.636	0.312	0.177	0.107	0.047	0.204
E8	0.634	0.230	0.256	0.086	0.050	0.185
E4	0.623	0.402	0.314	0.074	0.072	0.067
E1	0.613	0.415	0.255	0.126	0.271	－0.073
E3	0.606	0.496	0.180	0.135	0.158	－0.005
E10	0.596	0.398	0.228	0.073	0.146	－0.040
E5	0.590	0.334	0.210	0.256	0.185	0.012
E9	0.589	0.341	0.203	0.085	0.045	0.194
F2	0.584	0.312	0.307	0.111	0.304	0.069
F5	0.572	0.073	0.345	0.242	0.189	0.254
E2	0.566	0.527	0.199	0.121	0.229	－0.160
F8	0.564	0.115	0.265	0.204	0.205	0.095
F3	0.562	0.238	0.230	0.200	0.229	0.214
F7	0.551	0.308	0.232	0.078	0.080	0.355
F6	0.539	0.287	0.301	0.204	0.143	0.146
D10	0.514	0.260	0.383	0.138	0.276	－0.002
F9	0.507	0.202	0.305	0.024	0.171	0.353
D11	0.504	0.340	0.343	0.158	0.231	－0.116
F10	0.501	0.128	0.327	0.181	0.225	0.361
C2	0.350	0.747	0.081	0.118	0.145	0.144
C1	0.341	0.744	0.158	0.119	0.203	0.147
C5	0.296	0.736	0.243	0.135	0.214	0.060
C3	0.284	0.714	0.150	0.153	0.169	0.187
C4	0.268	0.692	0.196	0.256	0.186	0.064
C6	0.330	0.643	0.302	0.164	0.123	0.122
C7	0.370	0.634	0.211	0.206	0.112	0.147
B5	0.380	0.606	0.130	0.175	0.381	0.039
B8	0.234	0.495	0.173	0.101	0.266	0.315
B6	0.331	0.495	0.211	0.150	0.419	0.210
B7	0.246	0.453	0.276	0.072	0.413	0.372
D5	0.259	0.197	0.726	0.143	0.042	0.157
D7	0.358	0.239	0.652	0.109	0.184	0.020
D6	0.247	0.433	0.624	0.014	0.141	0.040
D4	0.359	0.127	0.623	0.232	0.078	0.136
D3	0.340	0.206	0.615	0.183	0.222	0.143
D2	0.291	0.253	0.601	0.199	0.172	0.181
D1	0.331	0.259	0.600	0.220	0.213	0.160
D9	0.306	0.118	0.570	0.238	0.231	0.060
D8	0.406	0.141	0.542	0.155	0.202	0.200
A7	0.103	0.053	0.217	0.840	0.083	－0.014
A6	0.249	0.046	0.212	0.753	0.170	0.039
A8	0.276	0.315	0.159	0.672	0.067	－0.025
A5	0.167	0.065	0.246	0.630	0.182	0.139
A4	0.170	0.248	－0.073	0.602	0.127	0.329
A3	0.133	0.156	0.008	0.543	－0.004	0.441
A9	0.149	0.489	0.153	0.506	0.062	0.126
A10	0.174	0.389	0.172	0.496	0.156	0.120
B4	0.242	0.204	0.143	0.198	0.773	0.091
B3	0.230	0.279	0.204	0.221	0.772	0.093
B2	0.236	0.288	0.250	0.120	0.702	0.138
B1	0.202	0.374	0.270	0.132	0.664	0.162
F12	0.365	0.058	0.418	0.111	0.123	0.539
A11	0.080	0.391	0.174	0.324	0.177	0.522
A12	0.162	0.257	0.226	0.307	0.256	0.494
F11	0.417	0.056	0.426	0.206	0.122	0.485

表5 第三次因子信度分析

因子名称 / 信度指标	管理与实施障碍	法律与资金障碍	技术障碍	认知障碍	成本障碍	用户障碍
克隆巴赫α系数	0.966	0.949	0.921	0.876	0.914	0.797
项目个数	26	11	9	8	4	4

表6 科学数据开放共享障碍分类表

障碍类型	障碍要素名称
管理与实施障碍	缺乏有效的科学数据开放共享政策;缺乏科学数据开放共享协调机制;缺乏科学数据开放共享管理机制;存在与科学数据开放共享相冲突的管理制度;存在限制科学数据开放共享的商业合同;缺少有效的科学数据开放共享计划;缺少科学数据开放共享途径;缺少与数据用户交互的经验;缺少处理用户反馈信息的环节;缺少共享数据的激励机制;缺少开放共享的组织文化;科学数据管理措施混乱;科学数据管理程序复杂;数据生产者和管理者之间缺乏互惠机制;数据提供者与数据用户之间缺乏信任;缺乏科学数据开放共享指南;数据提供者限制了数据的使用;机构内部或机构之间缺乏科学数据开放共享的各种合作;科学数据市场调节失灵;科学数据开放共享各国发展不平衡;开放共享科学数据在现有科学数据中所占比例太小;科学数据被垄断;缺乏数据资源开放共享领域的专业人才;数据用户缺乏判断或理解科学数据价值的能力;缺乏科学数据保护技术;缺乏科学数据开放共享标准
法律与资金障碍	缺乏完善的科学数据开放共享法律体系;存在科学数据开放共享安全问题;存在科学数据开放共享的隐私问题;容易失去科学数据所有权和版权的控制;存在科学数据的知识产权问题;存在科学数据的许可问题;可能面临科学数据开放共享的争议和诉讼风险;缺乏共享科学数据的报酬或奖励机制;缺乏足够的科学数据开放共享资金投入;需要增加预算外的研究成本;因科学数据开放共享而存在潜在的经济损失
技术障碍	存在科学数据完整性较差的问题;存在科学数据准确性较差的问题;存在科学数据及时性较差的问题;存在科学数据格式不标准问题;存在数据软件不兼容的问题;缺乏科学数据开放共享平台或系统;缺少科学数据共享的辅助工具(如索引、搜索引擎);缺乏元数据;存在国家间共享科学数据的语言障碍
认知障碍	研究者担心科学数据被别人误解;研究者担心科学数据被别人误用;研究者担心存在研究错误而受到批评;研究者担心打乱研究计划;研究者担心受到别人干扰;研究者担心失去对数据的控制;研究者担心泄露数据中的隐私信息;研究者害怕失去出版机会
成本障碍	搜集、提交和存储科学数据的时间成本高;搜集、提交和存储科学数据的人力成本高;提供、传递、维护和管理科学数据的时间成本高;提供、传递、维护和管理科学数据的人力成本高
用户障碍	数据用户缺少合理利用科学数据的知识;数据用户缺乏创建和保存元数据的知识;数据用户缺乏数据获取能力;数据用户缺乏数据处理能力

上述实证结论与第一部分理论分析观点在科学数据开放共享障碍类型及其要素上有些差异,这主要体现在如下几方面:①实证分析结果把原"管理障碍"与"实施障碍"综合起来,成为同一类障碍,即"管理与实施障碍"。这很可能是因为科学数据开放共享实践活动(即实施)往往离不开管理支持。不过,在此类障碍中,同时添加了"缺乏科学数据保护技术"和"缺乏科学数据开放共享标准"两项,而此两项原本被认为是"技术障碍"。产生此差异的原因可能是这两项选择在某些人看来就是管理不足问题(即管理障碍问题)。②由于法律和资金对于科学数据开放共享的支持都是必不可少的,实证分析结果把原"法律障碍"和部分"经济障碍"综合起来组成一类"法律与资金障碍"。③原"经济障碍"中的4个命题经因子分析后聚类成为"成本障碍"的4个选项,这说明人们很担心科学数据开放共享会造成许多不必要的成本。④原"认知障碍"和"实施障碍"中与数据用户相关的研究命题经因子分析后聚类成为"用户障碍"的4个选项,这说明成功的科学数据开放共享必须帮助用户解决其遇到的相关问题。

实证分析初步验证了科学数据开放共享确实存在许多障碍。这就启发我们今后应该针对上述障碍,制定有效的解决方案或对策,主要包括:①建立健全我国科学数据开放共享法律制度体系,加强数据立法工作,同时基于我国《科学数据管理办法》制定政府、机构、协会科学数据开放共享细则,如中国科学院发布的《中国科学院科学数据管理与开放共享办法(试行)》[27],使我国科学数据开放共享实践能够得到法律与政策的强力支持;②国家和相关研究机构、研究资助机构应该成立科学数据开放共享基金,为国家重大科学攻关项目、行业或机构重点研究项目的数据共享及其后续维护与利用提供充足的资金支持;③不断完善包括"国家基础科学数据共享服务平台""国家人口与健康科学数据共享服务平台""国家海洋科学数据共享服务平台""国家地球系统科学数据共享服务平台""国家标准文献共享服务平台""中国高等教育文献保障系统"和"国家科技图书文献中心"等在内的国家级科学数据共享平台建设,同时新建一批与国计民生紧密关联的国家级或行业级的科学数据开放共享平台,为机构和个人实施科学数据开放共享提供强大的技术支撑;④机构应该加强自身的科学数据共享平台建设,同时为用户提供有效的科学数据共享培训,努力解决研究

人员和用户科学数据开放共享过程中遇到的认知问题、成本问题和技术问题,尽量消除其在科学数据开放共享活动中的障碍,以便提高我国科学数据开放共享水平。

参考文献:

[1] OECD. Making open science a reality[R/OL].[2019-01-29]. http://wiki. lib. sun. ac. za/ images/0/02/Open-science-oecd. pdf.

[2] 熊易.国际开放科学数据实证资源及利用研究[J].图书馆理论与实践,2016(1):12-14.

[3] BJÖRK B. Open access to scientific publications - an analysis of the barriers to change? [EB/OL]. [2019-01-29]. http://www. informationr. net/ir/9-2/paper170. html.

[4] HEDRICK T E. Justifications for and obstacles to data sharing [M]//STEPHEN E F, MARGARET E M, MIRON L S. Sharing research data. Washington, DC: National Academy Press, 1985: 123-147.

[5] CHILDS S, MCLEOD J, LOMAS E, et al. Opening research data: issues and opportunities[J]. Records management journal, 2014, 24(2):142-162.

[6] MARTIN S, FOULONNEAU M, TURKI S, et al. Open data: barriers, risks, and opportunities[C]//FERRARI E, CASTELNOVO W. Proceedings of the 13th European conference on e-government. Como: Academic Conferences and Publishing International Limited, 2013:301-309.

[7] JANSSEN M, CHARALABIDIS Y, ZUIDERWIJK A. Benefits, adoption barriers and myths of open data and open government[J]. Information systems management, 2012, 29(4):258-268.

[8] BARRY E, BANNISTER F. Barriers to open data release: a view fromthe top[J]. Information polity, 2014, 19(1):129-152.

[9] STAGARS M. Open data in Southeast Asia[M]. Singapore: PalgraveMacmillan, 2016:17-20.

[10] SUI D. Opportunities and impediments for Open GIS[J]. Transactions in GIS, 2014, 18(1):1-24.

[11] PANHUIS W G V, PAUL P, EMERSON C, et al. A systematic review of barriers to data sharing in public health [EB/OL]. [2019-01-29]. https://bmcpublichealth. biomedcentral. com/articles/10. 1186/1471-2458-14-1144.

[12] VAN DEN EYNDEN V, KNIGHT G, VLAD A, et al. Towards open research: practices, experiences, barriers and opportunities [R/OL]. [2019-01-29]. https://ndownloader. figshare. com/files/6739038.

[13] ZUIDERWIJK A, JANSSEN M, CHOENNI S, et al. Socio-technical impediments of open data[J]. Electronic journal of e-government, 2012, 10(2):156-172.

[14] FIENBERG S E, MARTIN M E, STRAF M L. Sharing research data[J]. Medical care, 1986, 24(10):879-880.

[15] BORGMAN C L,青秀玲.科研数据共享的挑战[J].现代图书情报技术,2013(5):1-20.

[16] SAVAGE C J, VICKERS A J. Empirical study of data sharing by authors publishing in Plos journals [EB/OL]. [2019-01-29]. http://journals. plos. org/plosone/article? id = 10. 1371/journal. pone. 0007078.

[17] BORGMAN C L, WALLIS J C, ENYEDY N. Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries[J]. International journal on digital libraries, 2007, 7(1/2):17-30.

[18] SAYOGO D S, PARDO T A. Exploring the determinants of scientific data sharing: understanding the motivation to publish research data[J]. Government information quarterly, 2013, 30 (S1):19-31.

[19] TENOPIR C, ALLARD S, DOUGLASS K, et al. Data sharing by scientists: practices and perceptions [EB/OL]. [2018-10-9]. http://journals. plos. org/plosone/article? id = 10. 1371/journal. pone. 0021101.

[20] 黄如花,邱春艳.国外科学数据共享研究综述[J].情报资料工作,2013(4):25-31.

[21] PENG C, SONG X, JIANG H, et al. Towards a paradigm for open and free sharing of scientific data on global change science in China [EB/OL]. [2019-01-29]. https://esajournals. onlinelibrary. wiley. com/doi/epdf/10. 1002/ehs2. 1225.

[22] SEDRANSK N, YOUNG L J, KELNER K L, et al. Make research data public? not always so simple: a dialogue for statisticians and science editors[J]. Statistical science, 2010, 25(1):41-50.

[23] 扆铁梅,顾立平,董洁,等.国外科学数据开放获取研究[M].北京:中国财政经济出版社,2017:52

[24] 万望辉,崔辰州,乔翠兰,等.天文观测数据开放共享政策与策略分析研究[J].天文研究与技术,2015,12(3): 364-373.

[25] 陈传夫.中国科学数据公共获取机制:特点、障碍与优化的建议[J].中国软科学,2004(2):8-13.

[26] WESSELS B, FINN R L, LINDE P, et al. Issues in the development of open access to research data[J]. Prometheus, 2014, 32 (1):49-66.

[27] 中国科学院.《中国科学院科学数据管理与开放共享办法(试行)》[EB/OL]. [2019-02-21]. http://m. cas. cn/tzgg1/201902/P020190220358041915907. pdf.

作者贡献说明:

盛小平:负责论文撰写与修改;
吴红:参与问卷设计、问卷调查、数据处理与分析;
胡冰洁:参与问卷调查、资料收集。

图书情报工作
第 63 卷 第 17 期 2019 年 9 月

An Empirical Study on Barriers of Scientific Data Open Sharing

Sheng Xiaoping[1] Wu Hong[2] Hu Bingjie[2]
[1] School of Library, Information and Archives, Shanghai University, Shanghai, 200444
[2] School of Economics and Management, South China Normal University, Guangzhou 510006

Abstract: [**Purpose/significance**] This paper summarizes and verifies viewpoints of barriers of scientific data open sharing in order to promote open scientific theory research. [**Method/process**] Based on the theoretical analysis of barriers of scientific data open sharing, 64 research hypotheses are put forward and used as the questions of questionnaire. Through questionnaire survey and data processing and analysis, the relevant research hypotheses are verified. [**Result/conclusion**] The barriers of scientific data open sharing mainly include 26 kinds of managing and implementing barriers, 13 kinds of legal and financial barriers, 9 kinds of technical barriers, 8 kinds of cognitive barriers, 4 kinds of cost barriers and 4 kinds of user barriers. In order to improve the level of open sharing of scientific data in China, the relevant solutions or strategies for the above obstacles should be formulated.

Keywords: scientific data open sharing barrier empirical study